Abstract Title Page

Title: Can Comparison of Contrastive Examples Facilitate Graph Understanding?

Author(s): Linsey A. Smith, Dedre Gentner

Abstract Body

Background / Context:

Experts agree: comparison is good. Researchers in cognitive science emphasize the importance of comparison for learning and transfer (e.g., Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983). A large body of research demonstrates that comparison can lead learners to focus on deep relational commonalities rather than on specific, potentially idiosyncratic features of the particular examples (Gentner & Medina, 1998; Namy & Gentner, 2002; Ross & Kennedy, 1990). For example, when people are asked to discuss cigarettes, they write something like ‘‘cylindrical, paper-wrapped, filled with tobacco,’’ whereas when people compare cigarettes and time bombs, they write something like ‘‘They both do their damage after a period in which no damage is evident’’ (Gentner & Clement, 1988).

Experimental studies on comparison have yielded three key findings: (a) two examples are better than one (Gick & Holyoak, 1983; Namy & Gentner, 2002), (b) two examples presented together are better than two examples presented separately (Gentner et al., 2003; Oakes & Ribar, 2005), and (c) instructional support augments the benefits of comparison (Catrambone & Holyoak, 1989; Gentner et al., 2003; VanderStoep & Seifert, 1993). However, largely absent from this extensive cognitive science literature are investigations of the benefits of comparison with academic tasks, although recent research has looked at whether comparison can help children learn mathematical procedures (Rittle-Johnson and Star, 2007), or undergraduates learn complex geoscience concepts (Jee et al., 2010). In addition, the cognitive science literature provides limited guidance on one of the most important decisions that must be made in the implementation of comparison—namely, what should be compared. When two examples are to be compared, what dimensions of the examples should vary and what dimensions should remain the same? In this set of studies, we investigate whether comparison can help learning in another domain—graph representations—and what types of comparisons are most useful.

Purpose / Objective / Research Question / Focus of Study:

We explore the role of comparison in improving graph fluency. The ability to use graphs fluently is crucial for STEM achievement, but graphs are challenging to interpret and produce because they often involve integration of multiple variables, continuous change in variables over time, and omission of certain details in order to highlight central higher-order relations. Can comparison facilitate graph fluency by focusing learners on the relations between multiple variables? Furthermore, does the comparison of highly similar graphs facilitate performance to a greater degree than comparison less similar cases?

Setting:

This experimental research was conducted in the Language and Cognition Laboratory at Northwestern University.

Population / Participants / Subjects:

Experiment 1: 22 Northwestern undergraduates participated to fulfill a course requirement.
Experiment 2: 64 Northwestern undergraduates participated to fulfill a course requirement.

Intervention / Program / Practice:

Domain of Study

In these studies, we focus on learning about *Stock-and-Flow* graphs, which depict the relationship between a resource quantity (stock), and the inflows and outflows that alter them.

Any stock can be thought of as the amount of water in a tub. The water level accumulates the flow of water into the tub (the inflow) less the flow exiting through the drain (the outflow). The rate of change in the water level is the net flow, given by the difference between the inflow and outflow. As everyday experience suggests, the water level rises only when the inflow exceeds the outflow, falls only when the outflow exceeds the inflow, and remains the same only when the inflow equals the outflow. Prior work has shown that stock-and-flow graph problems are unintuitive and difficult, even in simple systems (like bathtubs) and even for highly educated people with strong technical backgrounds (Sterman & Booth Sweeney, 2002).

<u>Experiment 1</u>

Procedure. Experiment 1 took approximately 20 minutes to complete. Participants were randomly assigned to the *Comparison* or *Sequential* (Control) condition. Each participant received a packet which included several example graphs and target graphing problems. Participants were tested in individual booths. They were given as much time as they needed, which was roughly 20 minutes for both groups.

Training Materials. Three example Stock-and-Flow graphs were shown to participants. Each example consisted of (1) a graph showing inflows and outflows to a stock and (2) a graph with the corresponding stock level. Each example illustrated one of three key principles of stocks-and-flows:

- When inflow > outflow, stock is rising (I > O → S+)
- When inflow < outflow, stock is falling (I < O → S-)
- When inflow = outflow, stock is constant (I = O → S0)

We constructed two different sets of examples (Table 1). The three graphs within a set were highly similar—the only elements that differed were the inflow trajectory and the corollary stock graph. Participants only saw examples from one of the sets, not both.

(table 1 here)

Participants in both conditions were first introduced to stocks-and-flows in the context of CO_2 emissions, where the stock is the amount of atmospheric CO_2, inflow is CO_2 emissions, and outflow is CO_2 removal. Then participants were shown an example, with a brief description of what the graphs depicted. All participants then saw the three examples above and were asked to elaborate on the examples.

The only difference between the Sequential and Comparison groups involved how the example and elaboration task was structured. The Sequential group (n=11) saw all three examples presented on separate pages; the order was counterbalanced across participants. Below each example, participants were asked to elaborate on each example by describing "What is going on in the TOP (BOTTOM) graph?".

The Comparison group (n=11) was given two examples presented side-by-side. To ensure that the Comparison group saw all three examples, they were given two comparison sets. Thus one of the examples was shown twice; the repeated example was counterbalanced across participants. Three types of comparisons were possible: (1) I > O and I < O, (2) I > O and I = O, or (3) I < O and I = O. All participants received type (1) as the first comparison, then either saw type (2) or type (3); thus there were two possible comparison types, which were counterbalanced within the Sequential and Comparison groups. For each comparison set, participants were asked to list similarities and differences between the inflow/outflow graphs and stock graphs.

Test materials. After completing the training task, participants were given seven *graphical integration* problems (Sterman & Booth Sweeney, 2002). For each problem, the hypothetical behavior for two (out of the three) stock-and-flow variables is shown, and the participant must draw the missing variable's corresponding trajectory (Figure 1). The problems were presented in two orders—one the reverse of the other.

(figure 1 here)

<u>Experiment 2</u>

Procedure. The procedure was as in Experiment 1, however in this study participants were either assigned to the *High Similarity* (HiSim, n=32) condition or *Low Similarity* (LoSim, n=32) condition.

Materials. The materials were the same used in Experiment 1. For the HiSim comparisons, the examples were drawn from a single set of examples (mimicking the *Comparison* condition in Experiment 1). To create the LoSim comparisons, one example within each comparison pair was replaced with the equivalent example from the other set (e.g., Set A, I > O was replaced with Set B, I > O). Thus, while the underlying principles being compared were exactly the same in the HiSim and LoSim groups, the actual trajectories in the graphs were extremely dissimilar in the LoSim group. The target problems were as in Experiment 1.

Research Design:

Experiment 1 involved a 2 (Training Condition: Comparison vs. Sequential) x 2 (Problem Order) x 2 (Example Set) x 2 (Comparison Type) between-subjects factorial design. Experiment 2 was of the same design, except the two levels of *Training Condition* were HiSim and LoSim.

Data Collection and Analysis:

Participants used a pen to write out (and draw) their responses in the packet.

Scoring Target Problems. Two raters, blind to the hypotheses of the studies, scored the graphical responses for overall correctness. Each response was given a score of 1 or 0. These scores were summed for each participant (min = 0, max = 7). Inter-rater agreement across both studies was 95% (Cohen's kappa = .86). Conflicting scores were resolved through discussion.

Scoring Similarity and Difference Listings. Scoring is currently in-progress. We have designed a scoring rubric that aims to capture the quality of participants' similarity/difference listings. For example, an example of a high quality response involves descriptions that: integrate flows and stock, mention the important difference between graphs (e.g., that one graph shows I > O while the other shows I < O), and are conceptual rather than graph-bound. In contrast, a low quality response would: talk about all variables separately; make no mention of the important difference; use graph-bound rather than conceptual descriptions (e.g., referring to the inflow as *the blue line*); and focus on superficial features of the graphs (e.g., titles).

Analysis. For both experiments 1 and 2, a multivariate analysis of variance (ANOVA) was Summed score was entered as the dependent variable and Training Condition, Problem Order, Example Set, and Comparison Type were entered as independent variables. Additionally, once the similarity and difference listings are scored—to obtain a measure of comparison quality—we will examine the relationship between task performance and the quality of comparison. Our prediction is that participants who generate higher quality comparisons will have higher scores on the graph problems.

Findings / Results:

<u>Experiment 1</u>: Experiment 1 tested the hypothesis that comparing examples will facilitate performance on our graphical integration problems. There were no main effects or interactions involving Problem Order, Example Set, and Comparison Type, all *p*s > .21, so participants were

collapsed into two groups—Comparison and Sequential—for all further analyses. Overall, people in the Comparison condition had higher scores on the graph problems (M=5.00) than those in the Sequential Condition (M=3.12), $t(20)= 2.11$, $p<.05$, two-tailed.

Experiment 2: Experiment 2 tested the hypothesis that comparing high-similarity examples will facilitate performance to a greater degree that low-similarity comparisons. There were no main effects or interactions involving Problem Order, Example Set, and Comparison Type, all *p*s > .38, so participants were collapsed into two groups—HiSim and LoSim—for all further analyses. Participants in the HiSim condition had slightly higher scores (M=4.64) than participants in the LoSim condition (M=3.38), although this difference was not significant, $t(62) = 1.36$, $p=0.18$.

Conclusions:

These results suggest that comparison of highly similar examples promotes understanding and fluency with graphical representations of stock-and-flow scenarios, compared to a situation in which the same examples are not compared (Experiment 1). This finding supports prior work that demonstrates the benefits of comparing examples, and extends it to a novel domain—graphical representations. We suggest that pedagogical methods that assume that learners will abstract principles from single examples or that they will spontaneously draw comparisons across examples are likely to fall well short expectations. We suggest that one aim for instruction should be not simply to provide cases but to encourage active comparison of examples.

The results of Experiment 2 are somewhat less clear, but are still valuable for what they can tell us about the range of permitted variation between cases. Experiment 2 showed no difference in performance between those who compared high-similarity examples and low-similarity examples. One possible explanation for this finding is that the paired cases in the low-similarity condition were, in fact, quite similar to one another: while the trajectories of the flows and stocks may have differed, the perceptual features of the graphs were identical (e.g., the inflow line was always blue). Prior empirical work on comparison has shown that when two examples share surface features that are consistent with deeper relational commonalities, identifying these relational commonalities becomes much easier (consider a volleyball and a soccer ball vs. a volleyball and a football) (e.g., Gentner & Medina, 1998). In these studies, it is possible that the low-similarity examples were similar enough to facilitate high-quality comparisons, which in turn would lead to better learning. Once we finish coding participants' similarity and difference listings, we can assess whether both the high-similarity and low-similarity groups generated equally good comparisons.

We are not suggesting that comparison of cases is a cure-all. Even if learners do compare examples, they may be imperfect or incomplete in their identification of crucial relational commonalities between cases (Reeves & Weisberg, 1994). Completing our analysis of comparison quality and whether it is a good predictor of performance will give us insight into what types of comparisons are crucial for learning about graphical representations.

One limitation of our findings concerns the generalizability of the results. The current studies were conducted in a laboratory setting with undergraduates at an academically rigorous university. Whether we replicate these results with other populations in other settings is an open question. However, clarifying the conditions under which comparison can help or hurt learning of graphical representations in a highly controlled environment is an important first step in developing useful classroom interventions.

Appendices

Appendix A. References

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1147–1156.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393–405.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65, 263-297.

Namy, L. L,. & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology*: General, 131, 5-15.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T., Sageman, B., Ormand, C. J., & Tikoff, B. (2010). Analogical thinking in geoscience education. *Journal of Geoscience Education*, 58 (1), 2-13.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7,* 85–98.

Reeves, L. M., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin, 115*, 381–400.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99,* 561–574.

Ross, B. H., & Kennedy, P. K. (1990). Generalizing from the use of earlier examples in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 42–55.

Sterman, J. D. & Booth Sweeney, L. (2002). Cloudy Skies: Assessing Public Understanding of Global Warming. *System Dynamics Review, 18,* 36-54.

VanderStoep, S. W., & Seifert, C. M. (1993). Learning "how" versus learning "when": Improving transfer of problem-solving principles. *Journal of the Learning Sciences, 3,* 93–111.

Appendix B. Tables and Figures

Table 1. Example Sets used in Experiments 1 and 2.

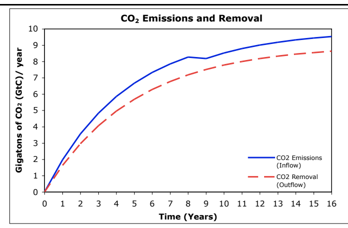
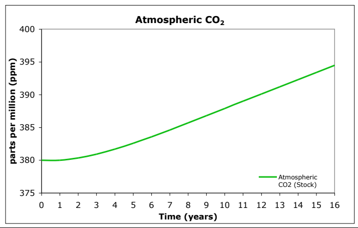
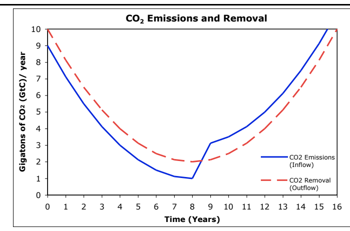
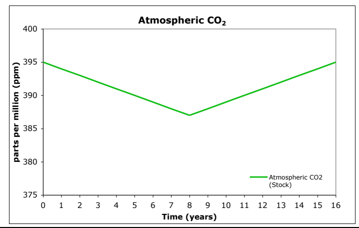
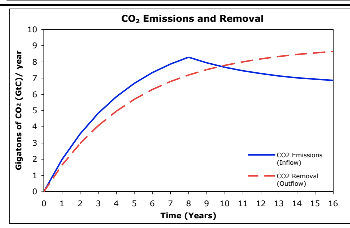
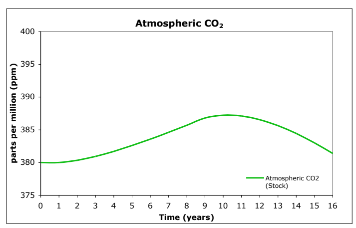
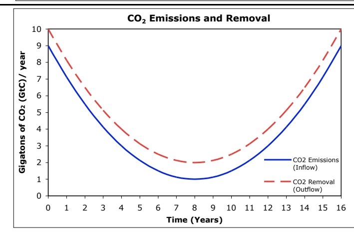
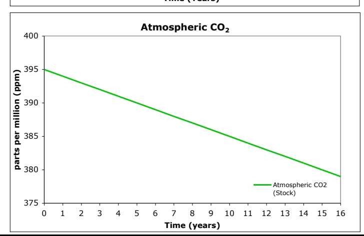
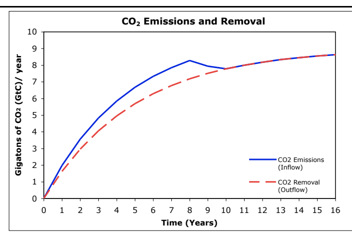
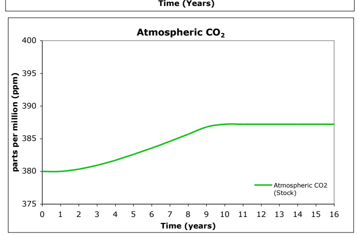
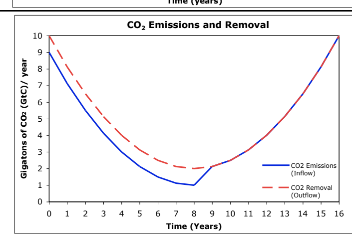
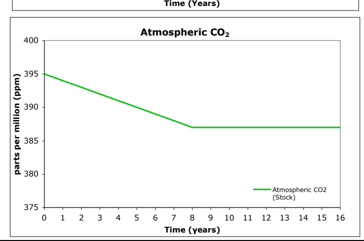
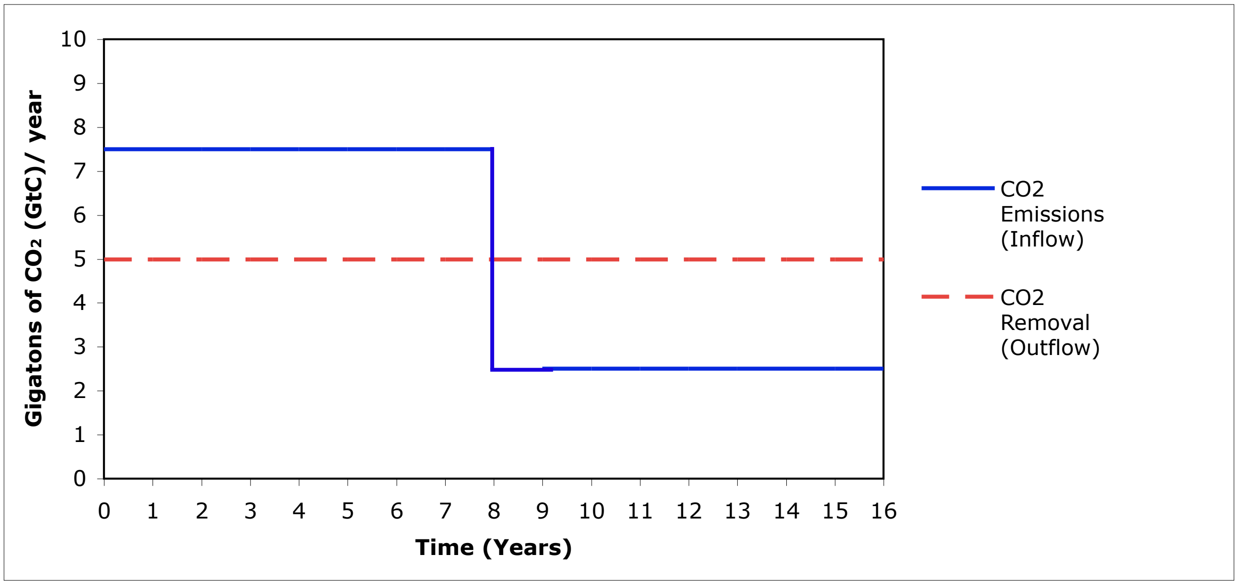
Key Principle Depicted	Set A	Set B
Inflow > Outflow, Stock Rising (I > O → S+)	 	 
Inflow < Outflow, Stock Falling (I < O → S-)	 	 
Inflow = Outflow, Stock Constant (I > O → S0)	 	 

Figure 1. Sample Graphical Integration Task. In this problem, the participant is given the inflow and outflow and must draw the corresponding stock.

Problem 1

The graph below shows a hypothetical pattern of CO_2 ***Emissions*** and ***Removal***.



On the graph below, draw the pattern of ***Atmospheric*** CO_2 that would be produced by the Emissions and Removal pattern above. The green dot (•) at time zero shows the initial atmospheric CO_2 level.